Thailand Statistician
April 2020; 18(2): 108-121
http://statassoc.or.th
Contributed paper

Improving Bivariate Ranked Set Sampling with Application to Chi-Square Control Chart

Panlop Nakakate [a], Sukuman Sarikavanij*[a], Saowanit Sukparungsee [b] and Yupaporn Areepong [b]

[a] Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi, Bangkok, Thailand.
[b] Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand.
*Corresponding author; e-mail: sukuman@hotmail.com, sukuman.sar@kmutt.ac.th

Received: 22 August 2018
Revised: 30 January 2019
Accepted: 16 March 2019

Abstract

Ranked set sampling (RSS) is useful for data collection if observations can be ranked cheaply without actual measurement. RSS has been studied widely but not enough, especially in case of collecting independent interested variables by using multivariate ranked set sampling. There are many problems in the sample selection for this case. One important problem is called "Incomplete Ranked Set Sampling (InRSS)", which causes bias in estimating mean. Hence, in this study, modified incomplete ranked set sampling (M-InRSS) is proposed for dealing with InRSS in one cycle bivariate RSS for independent bivariate normal distribution. Analytical results reveal that problems of bias and mean square error (MSE) in InRSS can be solved by using M-InRSS. Moreover, how to apply M-InRSS to chi-square control chart is shown with the numerical example of bivariate chi-square control chart based on M-InRSS. The results show that the charts based on M-InRSS are better than based on SRS.

Keywords: Incomplete ranked set sampling, Ridout's method of selecting sample, sample selection.

1. Introduction

Probability sampling has been used widely to collect data for a long time, so many sampling techniques are proposed for various conditions. One of them is ranked set sampling (RSS), which was proposed by McIntyre in 1952 (McIntyre 2005). It can be used for estimating mean by a sample whose units can be ranked without actual measurement. The RSS procedure is a two-stage-sampling (Chen et al. 2004). At the first stage, $n \times n$ sample units are collected by using simple random sampling (SRS) and divided into n equal sets. Then the sample units in each set are ranked separately (without actual measurement). At the second stage, one sample unit is selected from each set, and these selected sample units must have different ranks. Then the selected sample units are measured. This process is one cycle of RSS which can be repeated until enough sample units are collected. The number of cycle is denoted by r. Performing r cycle RSS will give $n \times r$ sample units. RSS is more effective than SRS in many applications of univariate problems (Chen and Wang 2004, Ozturk et al. 2005, Wang et al. 2009, Tiwari and Pandey 2013). However, only univariate statistics is not enough to solve problems in current researches. Hence, multivariate RSS has become a popular topic related to sampling for more than twenty years. One of the interesting issues in this topic is concerned with RSS on independent multivariate population. A problem in this issue is that the sample selection in the second stage is complicated. Ridout (2003) proposed a method of selecting samples to deal with this problem. It is a convenient method in which a sample could be selected easily. However, sample units according to some ranks might not be selected if the number of cycle is less than the number of variable. This situation is called "incomplete ranked set sampling (InRSS) whose arithmetic mean is a bias estimator (Nakakate and Sarikavanij 2017). Later, the Ridout method was modified to reduce InRSS occurrence by Nakakate et al. (2016). This method is called "modified Ridout ranked set sampling (M-RRSS)". However, the InRSS occurrence is not eliminated. Hence, in this study, we propose a method called "modified incomplete ranked set sampling (M-InRSS)" to improve M-RRSS under following assumptions. Population distribution is independent bivariate normal and means of order statistics are known.

In this study, we also demonstrated that M-InRSS could be applied to the chi-square control chart (Duncan 1950) based on RSS which was used to detect a shift of mean in a production process. This chart is consisted of center line (CL), upper control limit (UCL), and lower control limit (LCL). The effectiveness of this chart is considered by in-control average run length (ARL_0) and out-of-control average run length (ARL_1). These average run lengths (ARL) are used to measure "false alarm rate" and "detection ability", respectively. In application, if the sample used for constructing control limits (sample used for estimating mean and variance) is very large, this chart will be effective. Therefore, the population variance is reasonable to assume to be known in an application which is suitable for chi-square chart because an error in variance estimation is very small. This population variance can be used for determining the difference between two means of order statistics. This is the reason why the chi-square chart is selected to demonstrate the benefits of our proposed method. The independent assumption is selected due to the motivation from the following studies. The first study is to use the multivariate control chart for controlling the process of automotive body manufacturing in $X, Y,$ and Z dimensions. The results revealed that data of X and Z dimensions were independent (Liu et al. 2014). The latter is to use the control chart for controlling automotive stamped parts manufacturing by defining position of flange surface dimensions as characteristics. The results showed that covariance between SP9 and SP28 was equal to −0.0016 (Talib et al. 2014). We notice that two variables, which are independent, can be found in automotive industry. Therefore, our proposed method is an initial study for extending the bivariate control chart based on M-InRSS to trivariate cases with some correlated variables in the future.

The remaining parts of this article are arranged as follows. Univariate RSS is described in Section 2. Bivariate RSS is discussed in Section 3 consisting of four parts as follows. In Section 3.1, Ridout's method is reviewed. In Section 3.2, InRSS is reviewed, and then we determine mean square error (MSE) of mean estimator based on InRSS. In Section 3.3, M-RRSS is reviewed and discussed. In Section 3.4, we propose M-InRSS to improve M-RRSS, and then the expected value and MSE of mean estimator based on M-InRSS are determined. After that, the distribution shape of this mean is compared to normal probability density function (PDF) to check the suitability of chi-square test. In Section 4, we demonstrate how to apply M-InRSS to the bivariate chi-square control chart based on

RSS. In Section 5, the numerical example of the bivariate chi-square chart based on M-InRSS is shown, and Section 6 is the conclusion.

2. Ranked Set Sampling (RSS)

In this study, ranked set sampling (RSS) is selected for data collection. It was proposed first for a univariate case by McIntyre (2005). It is better than SRS for estimating mean. However, it is only suitable for a case in which orders of observations can be known without actual measurement. Variance of sample mean is reduced because of these orders. This reduction can be applied to reduce a sample size. Readers, who are interested in this issue, can study RSS procedures in Chen et al. (2004). The unbiased estimator of mean based on r cycle RSS with set size n is

$$\bar{X}_{RSS} = \frac{1}{nr}\sum_{j=1}^{r}\sum_{i=1}^{n} X_{[i:n]j},$$

when $X_{[i:n]j}$ is the i^{th} order observation of n sample from the j^{th} cycle, and r is a number of cycles (Wolfe 2012).

The variance of this mean estimator is as follows:

$$Var\left(\bar{X}_{RSS}\right) = E\left[\left(\bar{X}_{RSS}-\mu\right)^2\right] = \frac{1}{r^2n^2}\sum_{i=1}^{n} E\left[\left(X_{i:n}-\mu_{i:n}\right)^2\right] = \frac{1}{r^2}\left[\frac{\sigma^2}{n}-\frac{1}{n^2}\sum_{i=1}^{n}\left(\mu_{i:n}-\mu\right)^2\right],$$

when $\mu_{i:n}$ is the mean of the i^{th} order statistics in a sample of size n (Wolfe 2004).

This variance is clearly lower than the variance of mean estimator based on SRS. In this paper, only one cycle RSS ($r=1$) is considered, so a mean estimator based on one cycle RSS and its variance can be determined by substituting $r=1$ on the above equation.

3. Bivariate Ranked Set Sampling (BVRSS) and Problems

The bivariate ranked set sampling procedure is still similar to RSS. The mean estimator based on one cycle bivariate RSS ($\bar{\mathbf{X}}_{RSS}$) can be defined by using multivariate order statistic notation (Arnold et al. 2009) as

$$\bar{\mathbf{X}}_{RSS} = \frac{1}{n}\sum_{s=1}^{n}\mathbf{X}_{[\mathbf{i}_s:n]},$$

when $\mathbf{X}_{[\mathbf{i}_s:n]}$ is a two-dimensional-vector containing observation of all variables from ranks corresponding to vector $\mathbf{i}_s$, $\mathbf{i}_s$ is a two-dimensional-vector from the s^{th} set containing observation rank of all variables, respectively and n is set size.

A mean estimator for two interested variables based on RSS is a two-dimensional-vector containing all arithmetic means of variables. This estimator is unbiased so MSE of this estimator is as follows:

$$E\left(\left[\bar{\mathbf{X}}_{RSS}-\boldsymbol{\mu}_{RSS}\right]\left[\bar{\mathbf{X}}_{RSS}-\boldsymbol{\mu}_{RSS}\right]'\right) = \text{tr}\left(\boldsymbol{\Sigma}_{\bar{\mathbf{X}}_{RSS}}\right), \tag{1}$$

when $\boldsymbol{\mu}_{RSS}$ is a two-dimensional-vector containing population mean of both variables, $\boldsymbol{\Sigma}_{\bar{\mathbf{X}}_{RSS}}$ is a variance-covariance matrix of the mean estimator based on RSS, and tr(.) is a trace operator.

In addition, MSE of BVRSS is clearly lower than bivariate SRS. At the initial study about collecting samples for bivariate RSS, McIntyre RSS was applied to the bivariate rank set sampling with auxiliary variables in ordering (Norris et al. 1995). Although the method proposed by Norris works effectively for highly correlated variables, it is not suitable for our interested problem due to an independent assumption. In case of bivariate $\left(\mathbf{X}=\left[X_1 \quad X_2\right]\right)$, if X_1 and X_2 are independent, and the observations are ranked based on X_1, so selecting sample based on RSS for both variables will be complicated as shown in Figure 1. $X_{(i,i':n)s}$ in Figure 1 is denoted the sample units in the s^{th} set with ranked i and i' of the first and second variable, respectively. If sample selections are performed under the situation in this figure based on McIntyre, X_1 will be selected as RSS, but X_2 being order pair of X_1 will be selected as SRS.

Hence, the selecting sample based on RSS cannot be done regularly. As a matter of fact, there are other studies about sample selections for RSS which are suitable for this situation. In this paper, the method of selecting samples proposed by Ridout is modified for this situation.

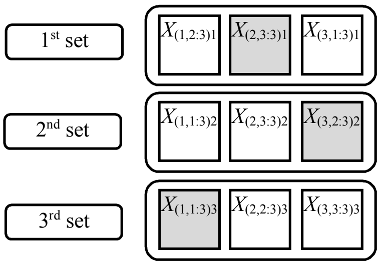

Figure 1 Selecting sample in one cycle RSS with set size three when X_1 and X_2 are independent

3.1. Ridout's method of selecting sample

After the first stage of RSS, ranked samples units are considered set by set in the second stage. Ridout chose the minimum sum of variance calculated by using sample unit quantity from each rank of each variable to be an indicator, which is used for selecting sample set by set. This sum of variance was defined by Ridout (2003) as follows:

$$V=\sum_{a=1}^{p}\sum_{i=1}^{n}\frac{\left(q_{a[i]}-\overline{q}_a\right)^2}{n-1}\ ,\ \overline{q}_a=\sum_{i=1}^{n}\frac{q_{a[i]}}{n},$$

where n is a set size, and $q_{a[i]}$ is a quantity of sample unit having the i^{th} order based on the a^{th} variable. This sum of variance can be rewritten in a vector form as

$$V=\frac{1}{n-1}\sum_{a=1}^{p}\mathbf{e}_a\mathbf{e}'_a,$$

where $\mathbf{e}_a=\left[q_{a[1]} \quad q_{a[2]} \quad \cdots \quad q_{a[n-1]} \quad q_{a[n]}\right]-\left[\overline{q}_a \quad \overline{q}_a \quad \cdots \quad \overline{q}_a \quad \overline{q}_a\right].$

Selecting sample by using this indicator is the same method as a sample selection that the sample units are selected under the condition. In this condition, the rank of sample unit must differ from the ranks of all previous selected sample units. If there are many sample units satisfying this condition, one of them will be selected randomly. Hence, the sample units in the first set are selected randomly, but the sample units in the last set are selected under many conditions from all of the previous sets. Although this method is convenient and fast, there are many cases which all sample units in the last set do not satisfy the conditions. For instance, if we have one cycle bivariate RSS with the set size three, the sample ranks are the same as shown in Figure 1, the selected sample units from the first and the second sets are from the 1st and the 2nd rank based on the first variable and the 2nd and the 3rd rank based on the second variable, respectively then either the sample unit on the left or on the right of the last set will be selected randomly because all sample units in the last set are at least one from the same rank of the first two. Hence, there is no sample unit from the 3rd rank based on the first variable or from the 1st rank based on the second variable in the sample set. This situation is called "Incomplete Ranked Set Sampling (InRSS)". Next, the effectiveness of the mean estimator based on InRSS is analyzed.

3.2. Incomplete Ranked Set Sampling (InRSS)

To consider one cycle BVRSS, InRSS will occur in only one variable from a sample based on one cycle RSS, which is collected by Ridout's method. The mean estimator for this variable (its arithmetic mean) is defined as

$$\bar{X}_{RSSmk} = \frac{1}{n}\left(\sum_{i\in M} X_{i:n} + X_{m:n}\right),\ m \in M,\ M = \{1,\ldots,n\} - \{k\},\ k \in \{1,\ldots,n\}, \tag{2}$$

where $X_{i:n}$ is the i^{th} order observation from all n orders, $X_{m:n}$ is the m^{th} order observation from all n orders, m is the repeated order, k is the missing order,

and
$$E(\bar{X}_{RSSmk}) = \mu + \frac{\mu_{m:n} - \mu_{k:n}}{n}. \tag{3}$$

Thus, it is obvious that the second term is the biased term. MSE of the mean estimator based on InRSS can be derived as follows:

$$E\left[\left(\bar{X}_{RSSmk} - \mu\right)^2\right] = E\left[\left(\frac{1}{n}\left(\sum_{i\in M} X_{i:n} + X_{m:n}\right) - \frac{1}{n}\sum_{i=1}^{n}\mu_{i:n}\right)^2\right] = E\left[\left(\frac{1}{n}\sum_{i\in M}\left(X_{i:n} - \mu_{i:n}\right) + \frac{1}{n}\left(X_{m:n} - \mu_{k:n}\right)\right)^2\right]$$

$$= E\left[\left(\frac{1}{n}\sum_{i\in M}\left(X_{i:n} - \mu_{i:n}\right)\right)^2\right] + 2E\left[\left(\frac{1}{n}\sum_{i\in M}\left(X_{i:n} - \mu_{i:n}\right)\right)\frac{1}{n}\left(X_{m:n} - \mu_{k:n}\right)\right] + E\left[\left(\frac{1}{n}\left(X_{m:n} - \mu_{k:n}\right)\right)^2\right]$$

$$= E\left[\left(\frac{1}{n}\sum_{i\in M}\left(X_{i:n} - \mu_{i:n}\right)\right)^2\right] + 2\left[E\left(\frac{1}{n}\sum_{i\in M}\left(X_{i:n} - \mu_{i:n}\right)\right)\right]\left[\frac{1}{n}E\left(X_{m:n} - \mu_{k:n}\right)\right] + E\left[\left(\frac{1}{n}\left(X_{m:n} - \mu_{k:n}\right)\right)^2\right].$$

Hence, we get

$$E\left[(\bar{X}_{RSSmk} - \mu)^2\right] = \frac{1}{n^2}\left(\sum_{i\in M} E\left[(X_{i:n} - \mu_{i:n})^2\right] + E\left[(X_{m:n} - \mu_{k:n})^2\right]\right). \tag{4}$$

The variance of the mean estimator based on RSS can be rewritten to the same form as follows:

$$E\left[\left(\bar{X}_{RSS} - \mu\right)^2\right] = \frac{1}{n^2}\sum_{i=1}^{n} E\left[\left(X_{i:n} - \mu_{i:n}\right)^2\right],$$

$$= \frac{1}{n^2}\left(\sum_{i\in M} E\left[\left(X_{i:n} - \mu_{i:n}\right)^2\right] + E\left[\left(X_{k:n} - \mu_{k:n}\right)^2\right]\right). \tag{5}$$

MSE of the mean estimator based on InRSS and RSS are compared by considering dissimilar terms in Equations (4) and (5), which are $E\left[\left(X_{m:n} - \mu_{k:n}\right)^2\right]$ and $E\left[\left(X_{k:n} - \mu_{k:n}\right)^2\right]$.

$$E\left[\left(X_{m:n}-\mu_{k:n}\right)^2\right] = E\left[\left(X_{m:n}-\mu_{\mathrm{m}:n}+\mu_{\mathrm{m}:n}-\mu_{k:n}\right)^2\right] = E\left[\left(X_{m:n}-\mu_{\mathrm{m}:n}\right)^2\right]+\left(\mu_{\mathrm{m}:n}-\mu_{k:n}\right)^2,$$

$$= Var(X_{m:n})+(\mu_{m:n}-\mu_{k:n})^2 = Var(X_{m:n})-Var(X_{k:n})+(\mu_{m:n}-\mu_{k:n})^2+Var(X_{k:n}).$$

In this study, we only consider the independent bivariate normal distribution. All possible cases of $Var(X_{m:n})-Var(X_{k:n})+(\mu_{m:n}-\mu_{k:n})^2+Var(X_{k:n})$ can be investigated by using the table of expected value of order statistics and products of order statistics (Teichroew 1956) when n starts from 2 to 20. The results from investigating reveal as follows.

In all cases, $Var(X_{m:n})-Var(X_{k:n})+(\mu_{m:n}-\mu_{k:n})^2>0.$

Hence, $E\left[(X_{m:n}-\mu_{k:n})^2\right]>E\left[(X_{k:n}-\mu_{k:n})^2\right]$ and $MSE(\bar{X}_{RSSmk})\geq MSE(\bar{X}_{RSS}).$

We can conclude that InRSS causes bias and the increase of MSE in estimating mean of the variable where InRSS occurs. Furthermore, InRSS increases overall MSE (Equation (1)). To deal with this problem, modified Ridout ranked set sampling (M-RRSS) (Nakakate et al. 2016) will be considered in Section 3.3. After that, modified incomplete ranked set sampling (M-InRSS) will be proposed in Section 3.4.

3.3. Modified Ridout Ranked Set Sampling

Modified Ridout ranked set sampling (M-RRSS) is a technique of selecting the sample modified from Ridout's method. It is different from Ridout's method because sample units are considered cycle by cycle. Therefore, if there are sample units in a cycle which satisfies a condition in Ridout's method of selecting sample, they will be selected. This selection reduces InRSS occurrence. Moreover, the indicator used for sample selection is modified. This modified indicator of M-RRSS for one cycle is defined as

$$\tilde{V}=\sum_{a=1}^{p}\tilde{\mathbf{e}}_a\tilde{\mathbf{e}}'_a,$$

where $\tilde{\mathbf{e}}_a=\begin{bmatrix}q_{a[1]} & q_{a[2]} & \cdots & q_{a[n-1]} & q_{a[n]}\end{bmatrix}-\begin{bmatrix}1 & 1 & \cdots & 1 & 1\end{bmatrix}$, and $q_{a[i]}$ is a quantity of the sample unit at the i^{th} order based on the a^{th} variable.

The second term of $\tilde{\mathbf{e}}_a$ is based on RSS, and it can be adjusted for many types of RSS modification. However, this study considers only the case of RSS. Nevertheless, this technique cannot eliminate InRSS occurrence in one cycle RSS. Hence, M-InRSS is proposed in this paper for solving the problem in bivariate cases. M-InRSS is applied when M-RRSS cannot eliminate InRSS occurrence. In the next section, the mean estimator based on M-InRSS will be defined, and then the expected value and MSE of this mean estimator will be analyzed. After that, the distribution of sample mean based on M-InRSS will be compared to normal PDF by using histograms.

3.4. Modified Incomplete Ranked Set Sampling (M-InRSS)

If M-RRSS is unsuccessful for selecting sample based on RSS, InRSS will occur and mean estimator will be biased as Equation (3). In case of normal distribution, this bias causes MSE to increase. We would like to modify the mean estimator based on InRSS to be unbiased. The value of sample unit, which is assigned to be the repeated order $(X_{m:n})$, will be added by the means of the m^{th} order statistics and negative value of the mean of k^{th} order statistics in this modification. This modified mean estimator of the improved method is called "M-InRSS". It is defined as follows.

$$\bar{X}_{M-InRSS} = \frac{1}{n}\left(\sum_{i\in M} X_{i:n} + X^{*}_{m:n}\right),$$

where $X^{*}_{m:n} = X_{m:n} - \mu_{m:n} + \mu_{k:n}$, and $X_{i:n}, m$ and $X_{m:n}$ are defined as in Equation (2).

Proposition 1 *Properties of a sample mean based on M-InRSS are as follows.*
(*a*) *The sample mean based on M-InRSS is unbiased.*
(*b*) *If* $m = n-k+1$, *the variance of the mean estimator based on M-InRSS will be equal to the variance of the mean estimator based on RSS.*

Proof:

(a) $$E\left(\bar{X}_{M-Inrss}\right) = \frac{1}{n}E\left(\sum_{i\in M} X_{i:n} + X^{*}_{m:n}\right) = \mu. \tag{6}$$

(b) Substitute $X_{m:n}$ by $X^{*}_{m:n}$ in Equation (4). We get the variance of this mean estimator as

$$E\left[\left(\bar{X}_{M-InRSS} - \mu\right)^2\right] = \frac{1}{n^2}\left(\sum_{i\in M} E\left[\left(X_{i:n} - \mu_{i:n}\right)^2\right] + E\left[\left(X_{m:n} - \mu_{m:n}\right)^2\right]\right). \tag{7}$$

For a symmetric distribution, $E\left[\left(X_{i:n} - \mu_{i:n}\right)^2\right] = E\left[\left(X_{n-i+1:n} - \mu_{n-i+1:n}\right)^2\right]$. Hence, if $m = n-k+1$, the variance of the mean estimator will become as

$$Var\left(\bar{X}_{M-InRSS}\right) = E\left[\left(\bar{X}_{M-InRSS} - \mu\right)^2\right] = E\left[\left(\bar{X}_{RSS} - \mu\right)^2\right] = Var\left(\bar{X}_{RSS}\right). \tag{8}$$

Equations (6) and (8) show that the mean estimator based on M-InRSS is unbiased and the variance of this mean estimator is the same as the variance of the mean estimator based on RSS. The result from modifying the estimator shows that the mean estimator based on M-InRSS in this case is not less effective than based on RSS, so this result is satisfied for estimating mean.

Although skewness of distribution of $X_{m:n}$ and $X_{k:n}$ are equal when $m = n-k+1$, they have opposite signs. Hence, the distribution of $\bar{X}_{M-InRSS}$ is slightly asymmetric. To illustrate the distribution shape of $\bar{X}_{M-InRSS}$, the set of selected order with set size n is defined as $\{o_1\ o_2\ \ldots\ o_n\}$. The element o_s is an order of selected sample unit from the s^{th} set. The permutations of the selected order are considered to be the same set. If a sample is selected based on one cycle McIntyre RSS, so the set of the selected order will be $\{1\ 2\ 3\ 4\ 5\}$. The histograms generated under M-InRSS with set sizes 3, 4, and 5 are compared to the normal PDF as shown in Figures 2 to 4. Those figures reveal that all histograms are close to the normal PDF.

Since RSS works effectively in univariate control charts (Abujiya and Lee 2013 and Al-Nasser et al. 2013), RSS is expected to work effectively in bivariate control charts. The next section will demonstrate how to apply M-InRSS in the chi-square control charts.

4. Chi-Square Control Chart

The chi-square control chart (Duncan 1950) is applied when there are more than one interested variable and known parameters. This section is divided into three parts which are control charts based on SRS, RSS and M-RRSS.

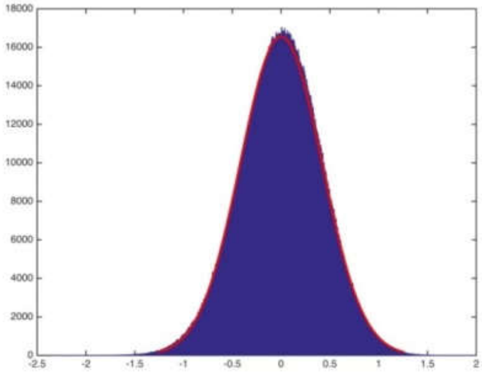

Figure 2 A normal curve fitting of sample mean based on M-InRSS with set size 3 and a sample set {1 2 1}

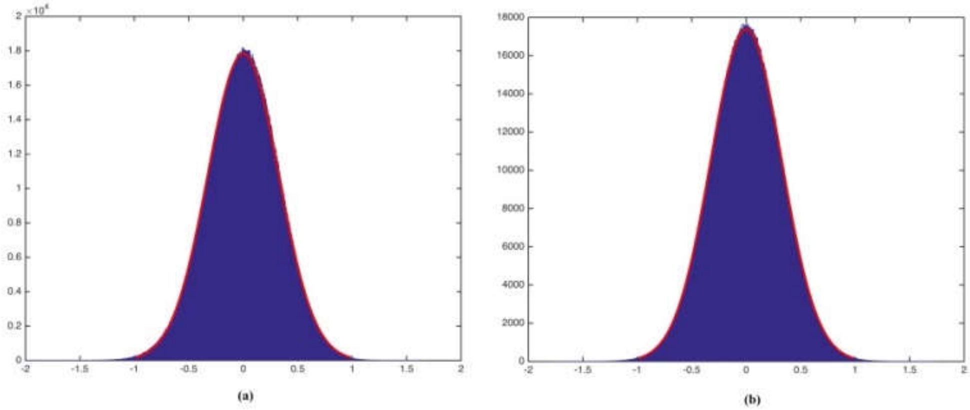

Figure 3 Normal curves fitting of sample mean based on M-InRSS with set size 4 and two sample sets; (a) {1 2 3 1}, (b) {1 2 2 4}

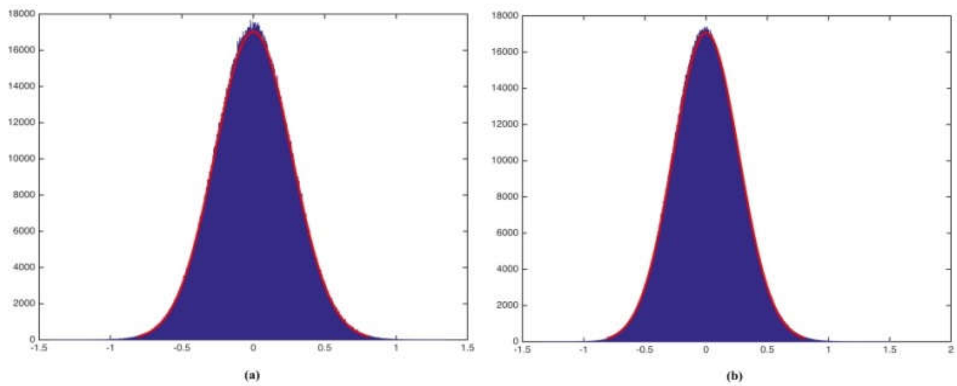

Figure 4 Normal curves fitting of sample mean based on M-InRSS with set size 5 and two sample sets; (a) {1 2 3 2 5}, (b) {1 2 3 4 1}

4.1. Chi-square control chart based on SRS

In general, a sample is collected based on SRS. If we have $\mathbf{x}=\left[x_1 \ldots x_p\right]$ and $\mathbf{x}$ is $\mathbf{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, statistics of this control chart based on SRS with sample size n (Rakitzis and Antzoulakos 2010) will be

$$T_i^2 = n\left[\overline{\mathbf{x}}_{SRS}-\boldsymbol{\mu}\right]'\boldsymbol{\Sigma}^{-1}\left[\overline{\mathbf{x}}_{SRS}-\boldsymbol{\mu}\right]=\left[\overline{\mathbf{x}}_{SRS}-\boldsymbol{\mu}\right]'\boldsymbol{\Sigma}_{\overline{\mathbf{x}}_{SRS}}^{-1}\left[\overline{\mathbf{x}}_{SRS}-\boldsymbol{\mu}\right], \tag{9}$$

when $\boldsymbol{\mu}$ is a $p\times 1$-dimensional-vector of population mean,

$\overline{\mathbf{x}}_{SRS}$ is a $p\times 1$-dimensional-vector of sample mean based on SRS,

$\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_{11} & \cdots & \sigma_{1p}\\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp}\end{bmatrix}$, $(\ .\ )^{-1}$ is an inverse matrix operator and

$$\boldsymbol{\Sigma}_{\overline{\mathbf{x}}_{SRS}}=\frac{1}{n}\Sigma,\ \ \Sigma=\begin{bmatrix}\frac{\sigma_{11}}{n} & \cdots & \frac{\sigma_{1p}}{n}\\ \vdots & \ddots & \vdots \\ \frac{\sigma_{p1}}{n} & \cdots & \frac{\sigma_{pp}}{n}\end{bmatrix}=\begin{bmatrix}\sigma_{\overline{\mathbf{x}}_{SRS(1,1)}} & \cdots & \sigma_{\overline{\mathbf{x}}_{SRS(1,p)}}\\ \vdots & \ddots & \vdots \\ \sigma_{\overline{\mathbf{x}}_{SRS(p,1)}} & \cdots & \sigma_{\overline{\mathbf{x}}_{SRS(p,p)}}\end{bmatrix}.$$

This statistic is distributed as χ_p^2, and the control limits of this chart are

$$LCL = CL = 0,$$

and

$$UCL=\chi^2_{\alpha,p}, \tag{10}$$

where $\chi^2_{\alpha,p}$ is the $(1-\alpha)$ percentile point of chi-square distribution with p degrees of freedom, and p is a number of the variables.

4.2. Chi-square control chart based on RSS

It is obvious that the variance term in Equation (9) depends on the sampling technique and the sample size. If a sample with size n is collected based on RSS, statistic of the chi-square control chart will be

$$T_i^2=\left[\overline{\mathbf{x}}_{RSS}-\boldsymbol{\mu}\right]'\boldsymbol{\Sigma}_{\overline{\mathbf{x}}_{RSS}}^{-1}\left[\overline{\mathbf{x}}_{RSS}-\boldsymbol{\mu}\right], \tag{11}$$

where $\overline{\mathbf{x}}_{RSS}$ is a $p\times 1$-dimensional vector of sample mean based on RSS, and

$$\boldsymbol{\Sigma}_{\overline{\mathbf{x}}_{RSS}}=\begin{bmatrix}\sigma_{\overline{\mathbf{x}}_{RSS(1,1)}} & \cdots & \sigma_{\overline{\mathbf{x}}_{RSS(1,p)}}\\ \vdots & \ddots & \vdots \\ \sigma_{\overline{\mathbf{x}}_{RSS(p,1)}} & \cdots & \sigma_{\overline{\mathbf{x}}_{RSS(p,p)}}\end{bmatrix}.$$

If we have $\mathbf{x}=\left[x_1 \ldots x_p\right]$ and $\mathbf{x}$ is $\mathbf{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\overline{\mathbf{x}}_{RSS}$ will be a multivariate normal distribution. Under assumptions of normality and perfect ranking in this paper, this statistic is also distributed as χ_p^2. Hence, the control limits of this chart are

$$LCL = CL = 0,$$

and

$$UCL=\chi^2_{\alpha,p}. \tag{12}$$

Since the sample selection in BVRSS is complicated, the technique of selecting samples like M-RRSS is necessary for the chi-square chart.

4.3. Chi-square control chart based on M-RRSS

The chi-square chart based on M-RRSS can be divided into two cases for consideration. If M-RRSS is succeeded to select a sample based on RSS, RSS will be applied in the chi-square chart. In contrast, M-InRSS will be applied in the chi-square chart if a selection based on M-RRSS is unsuccessful. From the comparison of histogram in Section 3.4, it is reasonable to set UCL of chi-square chart based on M-InRSS as the same as UCL of chi-square chart based on RSS because those histograms are close to normal PDF. Under known parameter assumption, $\mathbf{\Sigma}$ can be used to determine $\mu_{k:n}-\mu_{m:n}$ by using a mean of a standard normal order statistic table (Arnold et al. 1992) and scaling property of a normal distribution. Hence, $X^*_{m:n}$ can be determined from $X_{m:n}$. $\mu_{k:n}-\mu_{m:n}$ of the a^{th} variable can be determined as

$$\mu_{a(k:n)}-\mu_{a(m:n)}=\left(E\left[Z_{k:n}\right]-E\left[Z_{m:n}\right]\right)\cdot\sigma_{aa}=\left(E\left[Z_{k:n}\right]-E\left[Z_{m:n}\right]\right)\cdot\sigma_a^2,$$

where Z is $N(0,1)$, $\mu_{a(k:n)}$ is the mean of the k^{th} order statistics, and $\mu_{a(m:n)}$ is the mean of the m^{th} order statistics based on the a^{th} variable in a sample of size n. For example, if $\mathbf{X}$ is $\mathbf{N}_2(\boldsymbol{\mu},\mathbf{\Sigma})$, the set size equals 5, and InRSS occurs at X_2 with $k=2$ and $m=4$; then $\mu_{2(2:5)}-\mu_{2(4:5)}$ will equal $-0.990038\sigma_{22}$ or $-0.990038\sigma_2^2$. In addition, $\mathbf{\Sigma}_{\bar{\mathbf{x}}_{In-RSS}}=\mathbf{\Sigma}_{\bar{\mathbf{x}}_{RSS}}$ because σ_{12} and σ_{21} are zero.

For three charts from Sections 4.1 to 4.3, the stage of the process will be considered as the out-of-control stage and stopped in order to investigate some problems if $T_i^2 > UCL$, otherwise the process will be operated continually within the fixed time period, and then a new sample will be collected for reconsideration.

The effectiveness of the chi-square chart based on M-InRSS is illustrated in an example of the bivariate chi-square chart in the next section by using computer simulation. Both variables are assumed to be independent which makes all covariance in all cases become zero.

5. Numerical Results

From Section 3.2, InRSS causes the estimated mean to be biased. This bias causes type I error to increase in statistical hypothesis testing. Therefore, InRSS will cause the false alarm rate and the chi-square chart based on InRSS to be worse than the chi-square chart based on SRS. These reasons show that InRSS should not be applied in the control charts. Hence, we recommend that M-InRSS should be applied in the chi-square chart instead of InRSS. ARL of the chi-square control chart based on RSS and M-InRSS are used to be representatives of the chi-square control chart based on improved one cycle M-RRSS and are compared to ARL of the chi-square control chart based on SRS. These ARL are calculated from simulated data under the standard normal assumption. This comparison shows the effectiveness of the chi-square control chart based on M-InRSS. The numerical comparison of ARL is generated under the following assumptions. First, the bivariate standard distribution of population is normal, and both variables are independent. Second, the set sizes are 3, 4, and 5. These set sizes are chosen because they are used frequently, and the large set sizes often cause errors in ranking. Third, the shift sizes of both variables are denoted by δ_1 and δ_2, respectively, and they are set at 0.1, 0.5, 1.0, 1.5 and 2.0. The last assumption is concerned with the fact that the probability limits used for controlling ARL_0 of the chi-square chart based on RSS and SRS are equal. ARL_0 of the chi-square chart based on M-InRSS depends on the probability limit of the chi-square chart based on RSS according to the reason discussed in Section 4.3. In general, the probability limits are used in the case of non-normality population (Xie et al. 2002). However, a comparing detection ability by using ARL_1

should be controlled by setting ARL_0 of all cases equally. The probability limits work perfectly in this role. Parameters of the chi-square chart based on M-InRSS and RSS are equal because the mean estimators of these techniques are unbiased and the variances of these mean estimators are equal. The simulated ARL based on these sampling techniques are shown in Tables 1 to 3. The mirror images of the cases shown in these tables are omitted.

The comparison results from Tables 1 to 3 show that there are three main results. The first result shows that ARL_1 of the chi-square control chart based on RSS $(\text{ARL}_{1,\text{RSS}})$ is much lower than ARL_1 of the chi-square control chart based on SRS $(\text{ARL}_{1,\text{SRS}})$. Hence, the chi-square chart has better detection ability when it is performed based on RSS instead of SRS. The second result shows that ARL_1 of the chi-square control chart based on M-InRSS $(\text{ARL}_{1,\text{M-InRSS}})$ is slightly different from $\text{ARL}_{1,\text{RSS}}$ but much lower than $\text{ARL}_{1,\text{RSS}}$. Thus M-InRSS is suitable for solving InRSS problems. The last result shows that the increasing shift size causes $\text{ARL}_{1,\text{M-InRSS}}$ to be closer to $\text{ARL}_{1,\text{RSS}}$ in the shift of one and two variables. $\text{ARL}_{1,\text{M-InRSS}}$ and $\text{ARL}_{1,\text{RSS}}$ are not different in the following cases: 1) cases of {1 2 2 4}, {1 2 3 4 1} and {1 2 3 2 5} when $\delta_1 \geq 0$ and $\delta_2 \geq 0.3$, 2) cases of {1 2 3 1} when $\delta_1 \geq 0$ and $\delta_2 \geq 0.4$, and 3) cases of {1 2 1} when $\delta_1 \geq 0$ and $\delta_2 \geq 0.5$. From these results, we can conclude that the chi-square chart based on M-InRSS is effective in detecting the shift of mean when $\delta_1 \geq 0$ and $\delta_2 \geq 0.5$. Moreover, the differences of their ARL_1 are less than 0.03 at $\delta_1 \geq 0$ and $\delta_2 \geq 1$. Therefore, M-InRSS is effective in the chi-square chart because this chart is one of the Shewhart charts which is good for detecting the large shift size $(\delta \geq \sigma)$.

Table 1 The simulated ARL of chi-square chart with set size 3

Shift size		ARL		
δ_1	δ_2	SRS	RSS	M-InRSS {1 2 1}
0.0	0.0	370.3808	370.3736	365.6421
0.0	0.1	341.3636	319.4615	319.1775
0.0	0.2	270.4204	215.1017	220.0491
0.0	0.3	195.4385	130.5273	134.2454
0.0	0.4	135.1188	76.9693	78.5195
0.0	0.5	91.4463	45.6868	46.2957
0.0	1.0	15.2795	5.3980	5.3732
0.0	1.5	4.1083	1.6359	1.6437
0.0	2.0	1.7782	1.0783	1.0786
0.1	0.1	314.2497	275.8030	275.4929
0.2	0.2	207.9348	141.5608	144.7749
0.3	0.3	123.0312	67.5297	68.3031
0.4	0.4	70.8257	32.6996	32.9584
0.5	0.5	41.4042	16.8322	16.9000
1.0	1.0	4.8260	1.8912	1.8901
1.5	1.5	1.5469	1.0414	1.0426
2.0	2.0	1.0606	1.0004	1.0004

Table 2 The simulated ARL of chi-square chart with set size 4

Shift size		ARL			
δ_1	δ_2	SRS	RSS	M-InRSS	
				{1 2 3 1}	{1 2 2 4}
0.0	0.0	370.3597	370.3118	369.8751	366.4815
0.0	0.1	329.4120	289.6595	294.0902	289.5548
0.0	0.2	245.6153	164.5727	168.2725	163.9554
0.0	0.3	164.2360	84.5169	85.7630	83.9741
0.0	0.4	105.3943	43.3404	43.6854	43.2271
0.0	0.5	67.0061	22.9681	23.2032	23.0866
0.0	1.0	9.3439	2.4721	2.4843	2.48155
0.0	1.5	2.5512	1.1080	1.1167	1.1168
0.0	2.0	1.3323	1.0047	1.0030	1.0030
0.1	0.1	296.8657	234.4084	237.0478	234.0209
0.2	0.2	176.7981	94.5601	95.5047	94.3583
0.3	0.3	94.5478	36.9291	37.0874	36.9072
0.4	0.4	50.3040	15.7441	15.7529	15.7338
0.5	0.5	27.5963	7.5365	7.5496	7.5516
1.0	1.0	3.0706	1.1902	1.1885	1.1890
1.5	1.5	1.2140	1.0005	1.0009	1.0008
2.0	2.0	1.0096	1.0000	1.0000	1.0000

Table 3 The simulated ARL of chi-square chart with set size 5

Shift size		ARL			
δ_1	δ_2	SRS	RSS	M-InRSS	
				{1 2 3 4 1}	{1 2 3 2 5}
0.0	0.0	370.3656	370.3785	371.3230	372.7401
0.0	0.1	320.8103	262.6452	264.4636	262.6676
0.0	0.2	224.6287	123.6940	126.0813	124.2921
0.0	0.3	140.9219	54.7619	55.1087	54.7202
0.0	0.4	82.2780	25.1787	25.1925	25.1550
0.0	0.5	51.8697	12.5068	12.4977	12.5043
0.0	1.0	6.4337	1.5250	1.5262	1.5261
0.0	1.5	1.9057	1.0137	1.0131	1.0131
0.0	2.0	1.1407	1.0000	1.0001	1.0000
0.1	0.1	282.4474	197.2275	197.8654	196.2731
0.2	0.2	153.1623	62.8604	63.0087	62.9279
0.3	0.3	75.6020	21.0611	21.0350	21.0285
0.4	0.4	37.7944	8.3096	8.2980	8.2932
0.5	0.5	19.8878	3.9598	3.9552	3.9573
1.0	1.0	2.2283	1.0282	1.0284	1.0283
1.5	1.5	1.0799	1.0000	1.0000	1.0000
2.0	2.0	1.0014	1.0000	1.0000	1.0000

6. Conclusions

In this study, we propose a new method called M-InRSS for improving M-RRSS under bivariate normal assumption. This proposed method is one option to handle the InRSS problem which cannot receive the sample units from all ranks. The mean estimator based on M-InRSS is unbiased and has the same MSE as the mean estimator based on RSS. This method requires the difference between two means of order statistics. Applying M-InRSS in a control chart is studied by simulation under the independent bivariate standard normal assumption. ARL_0 and ARL_1 of the chi-square control chart in three cases which are SRS, RSS, and M-InRSS are compared. The results show that the false alarm rate and the detection ability of the M-InRSS case are close to the RSS case by considering ARL_0 and ARL_1, respectively. ARL_1 in the M-InRSS case is obviously lower than in the SRS case. Therefore, the effectiveness of the chart based on M-InRSS is almost the same as the chart based on RSS and better than the chart based on SRS. For the interesting issue in further study, it is possible to apply chi-square chart based on M-RRSS to the empirical data set which two variables are independent and the data collection is performed for a long period of time. For instance, data of automotive body manufacturing which is collected based on width, length, and height as $X, Y,$ and Z dimensions, respectively. From the study of Liu et al. (2014), X and Z dimensions are independent. However, using multivariate Shewhart charts based on M-RRSS for monitoring correlated variables has not been studied. Therefore, the multivariate chart based on M-RRSS for correlated variables is an interesting topic. In addition, applying this method to non-Shewhart multivariate control charts is another interesting topic for future researches.

Acknowledgements

This study cannot be done without valuable advice from members of Department of Mathematics, King Mongkut's University of Technology Thonburi, members of Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, and the late Professor Adisak Pongpullponsak.

References

Abujiya MR, Lee MH. The three statistical control charts using ranked set sampling. ICMSAO2013: Proceeding of the 5th International Conference on Modeling, Simulation and Applied Optimization; 2013 April 28-30; Hammamet, Tunisia. IEEE: 2013; 1-6.

Al-Nasser A, Al-Omari A, Al-Rawwash M. Monitoring the process mean based on quality control charts using on folded ranked set sampling. Pak J Stat Oper Res. 2013; 9(1): 79-91.

Arnold BC, Castillo E, Sarabia JM. On multivariate order statistics: application to ranked set sampling. Comput. Stat Data Anal. 2009; 53(12): 4555-4569.

Arnold BC, Nagaraja HN, Balakrishnan N. A first course in order statistics. New York: Wiley & Sons; 1992.

Chen Z, Wang YG. Efficient regression analysis with ranked-set sampling. Biometrics. 2004; 60(4): 997-1004.

Chen Z, Bai Z, Sinha BK. Ranked set sampling: theory and applications. New York: Springer-Verlag; 2004.

Duncan AJ. A Chi-square chart for controlling a set of percentages. Ind Qual Control. 1950; 71: 11-15.

Liu J, Pang CP, Jiang R. Economic optimization in automotive body manufacturing quality control based on multivariate Bayesian control chart. Int J Ind Syst Engrg. 2014; 18(1): 125-138.

McIntyre GA. A method for unbiased selective sampling, using ranked sets. Am Stat. 2005; 59(3): 230-232.

Nakakate P, Pongpullponsak A, Sarikavanij S. Modified Ridout's ranked set sampling on bivariate normal distribution. ICAS2016: Proceeding of International Conference on Applied Statistics 2016; 2016 July 13-15; Phuket, Thailand. 2016; 15-22.

Nakakate P, Sarikavanij S. Special case comparison between RSSMC and Ridout's ranked set sampling: incomplete order cases. Proceeding of the 98th The IIER International Conference; 2017 Mar 10; Pattaya, Thailand. IIER; 2017; 13-18.

Norris RC, Patil GP, Sinha AK. Estimation of multiple characteristics by ranked set sampling methods. Coenoses. 1995; 10(2-3): 95-111.

Ozturk O, Bilgin OC, Wolfe DA. Estimation of population mean and variance in flock management: a ranked set sampling approach in a finite population setting. J Stat Comput Simul. 2005; 75(11): 905-919.

Rakitzis AC, Antzoulakos DL. Chi-square control charts with runs rules. Methodol Comput Appl Prob. 2010; 13(4): 657-669.

Ridout MS. On ranked set sampling for multiple characteristics. Environ Ecol Stat. 2003; 10(2): 255-262.

Talib MA, Munisamy S, Ahmed S. Retrospective Hotelling's T^2 control chart for automotive stamped parts: a case study. J Sci Technol. 2014; 6(1): 101-113.

Teichroew D. Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution. Ann Math Stat. 1956; 27(2): 410-426.

Tiwari N, Pandey GS. Application of ranked set sampling design in environmental investigations for real data set. Thail Stat. 2013; 11(2): 173-184.

Wang YG, Ye Y, Milton DA. Efficient designs for sampling and subsampling in fisheries research based on ranked sets. ICES J Mar Sci. 2009; 66(5): 928-934.

Wolfe DA. Ranked set sampling: an approach to more efficient data collection. Stat Sci. 2004; 19(4): 636-643.

Wolfe DA. Ranked set sampling: its relevance and impact on statistical inference. ISRN Prob Stat. 2012; 2012: 1-32.

Xie M, Goh TN, Kuralmani V. Control charts with probability limits. Statistical models and control charts for high-quality processes. Boston: Springer; 2002; 21-38.